(19)日本国特許庁（ＪＰ）　(12)**公開特許公報（Ａ）**　(11)特許出願公開番号

特開平9－204376

(43)公開日　平成９年(1997)８月５日

(51)Int.Cl.[6]		識別記号	庁内整理番号	ＦＩ		技術表示箇所
Ｇ０６Ｆ	13/00	３５１		Ｇ０６Ｆ	13/00	３５１Ｃ
Ｈ０４Ｌ	12/46			Ｈ０４Ｌ	11/00	３１０Ｃ
	12/28		9466－5Ｋ		11/20	Ｂ
	12/66				13/00	３０５Ｚ
	29/06					

審査請求　未請求　請求項の数10　ＯＬ　（全 27 頁）

(21)出願番号　特願平8－11836

(22)出願日　平成８年(1996)１月26日

(71)出願人　000005223
富士通株式会社
神奈川県川崎市中原区上小田中４丁目１番１号

(72)発明者　飯塚　史之
神奈川県川崎市中原区上小田中1015番地
富士通株式会社内

(72)発明者　鳥居　悟
神奈川県川崎市中原区上小田中1015番地
富士通株式会社内

(74)代理人　弁理士　井桁　貞一

(54)【発明の名称】　通信宛先管理方法および装置

(57)【要約】

【課題】　本発明は多階層のネットワークプロトコルを用いたプロセス間通信技術に関し、ゲートウェイの負荷を低減し相手先のプロセッサに負荷を分散させること、および内部ネットワークに接続される並列／分散システムを１つのシステムとして扱うことを可能にする通信宛先管理方法および装置を提供することを目的とする。

【解決手段】　本発明通信宛先管理方法および装置は以下のように構成する。

（１）宛先ポートを読出し、対応する物理アドレスを書き換える。

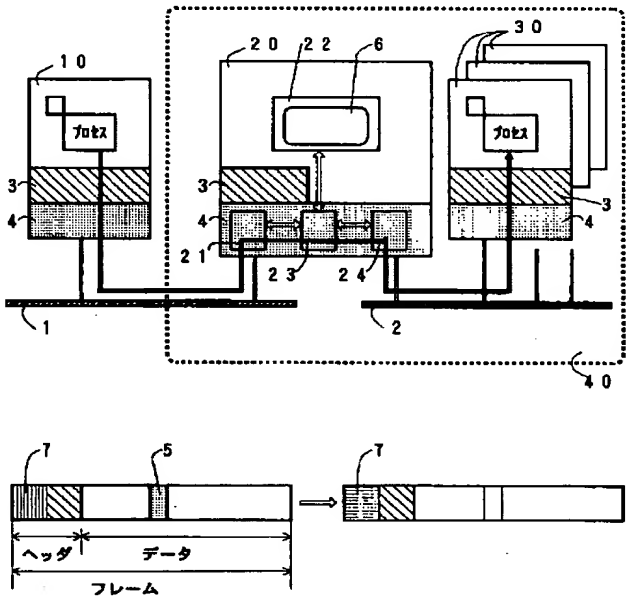
（２）ポートとプロセッサの物理アドレスの対応表記憶手段を設け、クライアントサーバ方式でポートを管理する。

（３）フラグメント記憶手段を設け、フラグメントの順序性を保証する。

（４）上記（１）と（３）を下位プロトコルで行うデータリンクブリッジ手段を設ける。

（５）複数のネットワークデータリンク処理手段を設ける。

第１の発明の原理図



1

【特許請求の範囲】

【請求項１】　外部のネットワークを介してホストコンピュータもしくはプロセッサエレメントもしくは他のゲートウェイに接続され、且つ、

内部のネットワークを介して複数のホストコンピュータもしくはプロセッサエレメントに接続されるゲートウェイを経由して、

上位プロトコルと下位プロトコルから成る階層構造のネットワークプロトコルによってプロセス間通信を行う場合、

通信の相手先プロセスに固有の前記上位プロトコルの宛先ポート情報を、前記下位プロトコルのフレームのデータ部分から読み出して、該宛先ポートから宛先の前記内部のホストコンピュータもしくはプロセッサエレメントのアドレスを調べ、前記下位プロトコルのフレームのヘッダ部分の宛先アドレスを書き換えることを特徴とする並列あるいは分散システムにおける通信宛先管理方法。

【請求項２】　内部のネットワークを介して複数のホストコンピュータもしくはプロセッサエレメントに接続されるゲートウェイにおいて、

前記ホストコンピュータもしくはプロセッサエレメントで実行されるプロセスがプロセス間通信で使用するポート情報と、該プロセスを実行する前記ホストコンピュータもしくはプロセッサエレメントのアドレスとの対応表を設け、

前記プロセスの実行開始時に、該プロセスの要求に応じて前記ポート情報を該プロセスに割り付け、該ポート情報を前記対応表に登録し、該プロセスの実行終了時に、該プロセスの要求に応じて前記対応表から該ポート情報を削除すると共に、

前記プロセスの要求に応じて、前記対応表を検索し前記ポート情報を該プロセスに通知することを特徴とする並列あるいは分散システムにおける通信宛先ポート管理方法。

【請求項３】　前記プロセス間通信中にパケットのフラグメント化が発生し、該通信の宛先ポート情報を含む先頭フラグメントが他のフラグメントより遅れて到着する場合に、

前記フラグメント化発生時に全フラグメントに付与される該パケットの識別子と共にフラグメントの順番を示す識別子を参照することによって、前記先頭フラグメントを特定し、該先頭フラグメントに含まれる前記下位プロトコルのフレームのヘッダ部分の宛先アドレスを書き換えてから全フラグメントを送信することを特徴とする請求項１に記載の並列あるいは分散システムにおけるフラグメントの通信宛先管理方法。

【請求項４】　前記ゲートウェイにおいて、前記下位プロトコルの内のデータリンク層に、

外部ネットワークとのインタフェースを持つ外部ネットワークデータリンク処理部と、内部ネットワークとのインタフェースを持つ内部ネットワークデータリンク処理部と、前記外部ネットワークデータリンク処理部と前記内部ネットワークデータリンク処理部の橋渡しをするデータリンクブリッジとを設け、

データリンクブリッジにより前記外部ネットワークからのパケットの宛先アドレスを書き換えて前記内部ネットワークに転送することを特徴とする請求項１～３のいずれかに記載の並列あるいは分散システムにおける通信宛先管理方法。

【請求項５】　前記ゲートウェイ内の前記データリンク層に、複数の前記内部データリンク処理部を設け、

複数の内部ネットワーク用のデータリンクを接続し、複数クラスタ内のプロセス間通信および複数クラスタと外部ネットワークとのプロセス間通信を行うことを特徴とする請求項４に記載の並列あるいは分散システムにおける通信宛先管理方法。

【請求項６】　上位プロトコルと下位プロトコルから成る階層構造のネットワークプロトコルによってプロセス間通信を行うためのゲートウェイであって、

外部のネットワークを介してホストコンピュータもしくはプロセッサエレメントもしくは他のゲートウェイを接続するためのインタフェースを持つ外部ネットワークデータリンク処理手段と、

内部のネットワークを介して複数のホストコンピュータもしくはプロセッサエレメントを接続するためのインタフェースを持つ内部ネットワークデータリンク処理手段と、

前記内部のネットワークに接続されるホストコンピュータもしくはプロセッサエレメントで実行されるプロセスがプロセス間通信で使用するポート情報と、該プロセスを実行する前記内部のホストコンピュータもしくはプロセッサエレメントのアドレスとの対応表を格納しておく対応表記憶手段と、

前記対応表記憶手段、前記外部ネットワークデータリンク処理手段、および前記内部ネットワークデータリンク手段に接続され、前記プロセス間通信の宛先を管理する宛先管理手段とを備え、

前記外部ネットワークデータリンク処理手段によって得られる通信の相手先プロセスに固有の前記上位プロトコルの宛先ポート情報を、前記下位プロトコルのフレームのデータ部分から読み出し、

前記記憶手段に格納された前記対応表から前記宛先ポートに対応する宛先のホストコンピュータもしくはプロセッサエレメントのアドレスを読み出し、

前記下位プロトコルのフレームのヘッダ部分の宛先アドレスを書き換え、

前記内部ネットワークデータリンク手段を経由して前記内部のホストコンピュータもしくはプロセッサエレメントに送信することを特徴とする並列あるいは分散システムにおける通信宛先管理装置。

3

【請求項７】　内部のネットワークを介して複数のホストコンピュータもしくはプロセッサエレメントに接続されるゲートウェイであって、
前記ホストコンピュータもしくはプロセッサエレメントで実行されるプロセスがプロセス間通信で使用するポート情報と、該プロセスを実行する前記ホストコンピュータもしくはプロセッサエレメントのアドレスとの対応表を格納しておく対応表記憶手段と、
前記ホストコンピュータもしくはプロセッサエレメントに備えられたポート管理クライアント手段を介して前記プロセスから発生する要求を処理するポート管理サーバ手段とを備え、
前記ポート管理サーバ手段によって、前記ポート情報を前記プロセスに割り付け、該ポート情報を前記対応表記憶手段に登録し、前記対応表記憶手段から該ポート情報を削除すると共に、
前記プロセスの要求に応じて前記対応表記憶手段を検索し前記ポート情報を該プロセスに通知することを特徴とする並列あるいは分散システムにおける通信宛先管理装置。

【請求項８】　前記対応表記憶手段、前記外部ネットワークデータリンク処理手段、および前記内部ネットワークデータリンク手段に接続され、フラグメントを格納するフラグメント記憶手段を備えたフラグメント処理手段を備え、
前記プロセス間通信中にパケットのフラグメント化が発生し、該通信の宛先ポート情報を含む先頭フラグメントが、他のフラグメントより遅れて到着する場合に、
前記フラグメント化発生時に全フラグメントに付与される該パケットの識別子と共にフラグメントの順番を示す識別子を参照することによって、前記先頭フラグメントを特定し、該先頭フラグメントに含まれる前記下位プロトコルのフレームのヘッダ部分の宛先アドレスを書き換えてから全フラグメントを送信することを特徴とする請求項６に記載の並列あるいは分散システムにおけるフラグメントの通信宛先管理装置。

【請求項９】　前記下位プロトコルの内のデータリンク層に、
前記外部ネットワークデータリンク処理手段と、前記内部ネットワークデータリンク処理手段と、
前記宛先管理手段および前記フラグメント処理手段とから構成されるデータリンクブリッジ手段とを備え、
前記データリンク層で、前記外部ネットワークからのパケットを該宛先アドレスに応じて前記内部ネットワークに転送することを特徴とする請求項６～８のいずれかに記載の並列あるいは分散システムにおける通信宛先管理装置。

【請求項１０】　前記データリンク層に複数の前記内部ネットワークデータリンク処理手段を備え、
複数の内部ネットワークを接続し、外部ネットワークとのプロセス間通信を行うことを特徴とする請求項９に記載の並列あるいは分散システムにおける通信宛先管理装置。

4

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は並列システムあるいは分散システムにおける通信宛先管理方法および装置に関わり、特に多階層のネットワークプロトコルを用いたプロセス間通信技術に関する。

【０００２】

【従来の技術】近年、大規模な分散システムの開発が本格化してきており、大量のデータ処理を高速に行う並列システムの開発や、クライアントサーバ方式の様々なアプリケーションの増加に伴い、プロセス間通信を伴うメッセージ連携のソフトウェア開発が盛んになってきている。

【０００３】メッセージ通信を効率的に行う研究用システムの開発や、業務用システムにおける並列システムや様々なクライアントサーバシステムの導入、更にマルチベンダーの異機種間接続などが進んでおり、速度が速くて効率良く使えるプロセス間通信が望まれている。

【０００４】こうしたプロセス間通信では、並列システムの内部に多数のプロセッサエレメントやホストコンピュータを備え、ローカルエリアネットワーク（ＬＡＮ）などの内部ネットワークを介して接続し、ワイドエリアネットワーク（ＷＡＮ）やＬＡＮなどの外部ネットワークを介して他の並列システムやホストコンピュータを接続し、且つそれぞれのネットワークが階層構造をしたプロトコルを持つネットワーク間通信を行う。

【０００５】階層化プロトコルには、一般に二つのモデルが知られている。一つは図１７に示すＩＳＯ（国際標準化機構）の７層参照モデルであり、１層の物理ハードウェア接続から７層のアプリケーションまでの７層の概念層から成る。もう一つは図１８に示すＴＣＰ／ＩＰＩｎｔｅｒｎｅｔ階層モデルであり、最下層のハードウェア層およびネットワークインタフェース層からアプリケーションまでの４層の概念層（合計５層）から成る。（ＴＣＰ／ＩＰによるネットワーク構築　８２ｐ～８５ｐ　共立出版　１９９０年７月）

本発明は、これらのモデルの階層間で渡されるオブジェクトをパケットとして転送するときの宛先アドレスに関わる。宛先アドレスはＴＣＰ／ＩＰ（Ｔｒａｎｓｍｉｓｓｉｏｎ　Ｃｏｎｔｒｏｌ　Ｐｒｏｔｏｃｏｌ／Ｉｎｔｅｒｎｅｔ　Ｐｒｏｔｏｃｏｌ）の通信プロトコルにおいては、ＴＣＰやＵＤＰ（Ｕｓｅｒ　Ｄａｔａｇｒａｍ　Ｐｒｏｔｏｃｏｌ）のトランスポート層などの上位プロトコルでは宛先ポート情報として、ＩＰ（Ｉｎｔｅｒｎｅｔ　Ｐｒｏｔｏｃｏｌ）層などの中位層プロトコルではＩＰ（Ｉｎｔｅｒ　Ｐｒｏｃｅｓｓ）アドレスとして、データリンク層などの下位プロトコルではホストコ

ンピュータもしくはプロセッサエレメントの（物理）アドレスとして定義される。一方、外部ネットワークに接続し通信を行うための機器として、ＩＰルータやブリッジやゲートウェイがあり、ネットワークとネットワークを接続している。またハブは１台で複数のネットワークを接続することができ、複数のクラスタ接続を可能にしている。

【０００６】従来のプロセス間通信では、図１９に示すようにＩＰルータやブリッジを用いてＩＰアドレスによりパケット転送をするものや、ゲートウェイ上にネットワークサーバを設けてパケット転送をするものが知られている。

【０００７】

【発明が解決しようとする課題】従来のＩＰルータやブリッジにおいては、システムを構成する全てのプロセッサエレメントにＩＰアドレスを付与する必要があり、プロセスのユーザも目的のプロセスがどのＩＰアドレスのプロセッサで実行されているかを認識しなければならないという問題があった。

【０００８】並列システムや分散システムが大規模になるに従いユーザが目的のプロセスとＩＰアドレスの対応を認識することは困難になり、またシステムの大規模化に伴いＩＰアドレスが不足することが予想される。

【０００９】一方、ゲートウェイ上でネットワークサーバを動作させる場合は、外部ネットワークデータリンク層、ネットワーク層、トランスポート層の各層のプロトコル処理を順次行ってから宛先ポートの情報を得るために、ゲートウェイの負荷が増大するという問題があった。

【００１０】更に内部ネットワークに向けてプロセス間通信用パケットを送出するためには、内部ネットワーク用プロトコル処理であるチェックサムを取るとか、データコピーをするなどの処理ステップのために通信速度が遅くなってしまうという問題があり、システムの規模が大きくなればそれだけゲートウェイに負荷が集中してしまうという問題があった。

【００１１】本発明はこのような点にかんがみてゲートウェイの負荷を低減し通信の相手先のプロセッサに負荷を分散させること、および内部ネットワークに１つのプロセッサアドレスを割り当てて並列システムもしくは分散システムを１つのシステムとして扱うことを可能にすると共に、複数クラスタの通信宛先管理を可能とし更に大規模な並列システムもしくは分散システムを構築するという通信宛先管理方法および装置を提供することを目的とする。

【００１２】

【課題を解決するための手段】上記の課題は下記の如くに構成された本発明の並列あるいは分散システムにおける通信宛先管理方法および装置によって解決される。

【００１３】（１）本発明の第１は、請求項１および請求項６にそれぞれ記載の通信宛先管理方法および装置である。図１は第１の発明の原理図である。図１において１は外部ネットワーク、２は内部ネットワーク、３は上位プロトコル、４は下位プロトコルを示す。上位プロトコル３と下位プロトコル４の中間には一般にＩＰプロトコルなどの中位層プロトコルがあるが図示していない。図のフレームの中にあるデータ部分にある５は宛先ポート情報、同じくヘッダ部分にある７は宛先アドレスを示す。

【００１４】１０は外部ネットワーク１に接続されるホストコンピュータもしくはプロセッサエレメントである。１０は図示していない他のゲートウェイを経由して外部ネットワークに接続される場合もある。２０はゲートウェイであり、本発明の通信宛先管理装置である。

【００１５】３０は内部ネットワーク２に接続されるホストコンピュータもしくはプロセッサエレメントであって複数ある。点線で囲った４０は一つの並列システムあるいは分散システムである。

【００１６】６は宛先ポート情報５と宛先アドレス７の対応を示すポート情報に関するアドレス対応表である。ポート・アドレス対応表６は、内部ネットワーク２に接続されるホストコンピュータもしくはプロセッサエレメントで実行されるプロセスがプロセス間通信で使用するポート情報と、該プロセスを実行する前記内部のホストコンピュータもしくはプロセッサエレメントのアドレスとの対応を示す。

【００１７】２１は外部ネットワーク１を接続するためのインタフェースを持つ外部ネットワークデータリンク処理手段、２２はポート・アドレス対応表６を格納する対応表記憶手段、２４は内部ネットワーク２を接続するためのインタフェースを持つ内部ネットワークデータリンク処理手段である。

【００１８】２３は外部ネットワークデータリンク処理手段２１、対応表記憶手段２２、および内部ネットワークデータリンク手段２４に接続され、プロセス間通信の宛先を管理する宛先管理手段である。

【００１９】第１の発明の方法はゲートウェイである第１の発明の装置２０を経由して、上位プロトコル３と下位プロトコル４から成る階層構造のネットワークプロトコルによってプロセス間通信を行う場合に、通信の相手先プロセスに固有の上位プロトコルの宛先ポート情報５を、下位プロトコル４のフレームのデータ部分から読み出して、宛先ポートから宛先の内部のホストコンピュータもしくはプロセッサエレメント３０のアドレスを調べ、下位プロトコルのフレームのヘッダ部分の宛先アドレス７を書き換えることを特徴とする。

【００２０】第１の発明の装置２０は、前記第１の発明の方法に記載のプロセス間通信を行うためのゲートウェイであって、外部ネットワークデータリンク処理手段２１と、対応表記憶手段２２と、宛先管理手段２３と、内

部ネットワークデータリンク処理手段２４とを備える。

【００２１】宛先管理手段２３では、まず相手先プロセスの上位プロトコル３の宛先ポート情報５を読み出す。宛先ポート情報５は外部ネットワークデータリンク処理手段２１によって得られる、下位プロトコル４のフレームのデータ部分にある。

【００２２】宛先管理手段２３では、次に対応表記憶手段２２に格納された対応表６から宛先ポート情報５に対応する宛先のホストコンピュータもしくはプロセッサエレメントのアドレス７を読み出す。

【００２３】読み出された宛先アドレス７が下位プロトコル４のフレームのヘッダ部分の宛先アドレスに書き換えられる。続いて下位プロトコル４のフレームが内部ネットワークデータリンク手段２４を経由して内部のホストコンピュータもしくはプロセッサエレメント３０に送信される。

【００２４】第１の発明によって、該ゲートウェイに接続される内部ネットワークを介して接続されるホストコンピュータやプロセッサエレメントのポート情報と該アドレスを管理しておけば、該ゲートウェイに一つのＩＰアドレスを割り付けるだけで内部ネットワークでのプロセス間通信および外部ネットワークとのプロセス間の通信が可能になると共に、上位プロトコルによる処理を実行しなくて済み該ゲートウェイの負荷を低減することが可能になる。

【００２５】（２）本発明の第２は、請求項２および請求項７にそれぞれ記載の通信宛先管理方法および装置であり、前記第１の発明で使用される対応表に関するものである。図２は第２の発明の原理図である。図２の２５は内部ネットワークに接続されるホストコンピュータもしくはプロセッサエレメント３０に備えられたポート管理クライアントに対応するポート管理サーバである。図２のその他の符号は図１と同じである。

【００２６】ポート情報に関するアドレス対応表６には、内部のホストコンピュータもしくはプロセッサエレメント３０で実行されるプロセスについてのポート情報と、そのプロセスが実行されるホストコンピュータもしくはプロセッサエレメント３０のアドレスが示される。

【００２７】第２の発明の方法は、内部のネットワークを介して複数のホストコンピュータもしくはプロセッサエレメント３０に接続されるゲートウェイ２０において、ホストコンピュータもしくはプロセッサエレメント３０で実行されるプロセスがプロセス間通信で使用するポート情報５と、該プロセスを実行する前記ホストコンピュータもしくはプロセッサエレメント３０のアドレスとの対応表６を設け、対応表６を検索すること、およびプロセスの要求に応じて、ポート情報をプロセスに割り付けること、対応表６に登録すること、および対応表６から削除することを特徴とする。

【００２８】第２の発明の装置は前記第２の発明の方法に記載のゲートウェイであって、対応表記憶手段２２とポート管理サーバ手段２５とを備え、前記第２の発明の方法を実現する。

【００２９】一つの並列システムあるいは分散システム４０の中のプロセスから、プロセスの実行開始時にポートの割付要求が発行されると、ポート管理クライアントからポート管理サーバ２５にそのポートの発行および登録を依頼する。ポート管理サーバ２５は、対応表記憶手段２２に格納されているポート・アドレス対応表６を参照し、未使用のポートを選び、その対応関係を対応表６に登録すると共に選ばれたポートをポート管理クライアントに通知する。ポート管理クライアントはプロセスにそのポートを割り付ける。

【００３０】ポートを開放する場合や、プロセスの終了時には、ポート・アドレス対応表６のエントリはポート管理サーバ２５によって削除される。またポート管理サーバ２５はプロセスの要求に応じて対応表を検索しポート情報をプロセスに通知することもある。

【００３１】第２の発明によって対応表６を常に更新することができ、また検索することができるので前記第１の発明による通信宛先管理を維持することが可能になる。

（３）本発明の第３は、請求項３および請求項８にそれぞれ記載の通信宛先管理方法および装置であり、前記第１の発明のプロセス間通信で使用されるパケットのフラグメント化が発生した場合に関するものである。図３は第３の発明の原理図である。

【００３２】フラグメントは、上位層パケットから中位層データグラムへの変換の際に、物理フレームに載るように中位層データグラムを分解するときに発生する。フラグメント化発生時には、該パケットの識別子と共にフラグメントの順番を示す識別子が全フラグメントに付与される。

【００３３】また物理層の中には転送の順序性が保証されていない物理層があり、これを使用する場合は宛先ポートを含む先頭のフラグメントが先頭以外のフラグメントより遅れて到着することがある。

【００３４】図３の２６がフラグメントを格納するフラグメント記憶手段であり、２７がフラグメント記憶手段２６を備えたフラグメント処理手段である。フラグメント処理手段２７は、対応表記憶手段２２、外部ネットワークデータリンク処理手段２１、および内部ネットワークデータリンク手段２４に接続される。図３のその他の符号は図１と同じである。

【００３５】第３の発明の方法は、プロセス間通信中にパケットのフラグメント化が発生し該通信の宛先ポート情報を含む先頭フラグメントが他のフラグメントより遅れて到着する場合に、ゲートウェイである通信宛先管理装置２０においてフラグメント化発生時に付与されるパケット識別子および順番識別子を参照することによっ

て、先頭フラグメントを特定し先頭フラグメントに含まれる宛先アドレス７を書き換えてから全フラグメントを送信することを特徴とする。

【００３６】第３の発明の装置は前記第３の発明の方法に記載のゲートウェイであって、フラグメント処理手段２７を備え、前記第３の発明の方法を実現する。第３の発明によって、パケット通信の過程でフラグメントが発生し先頭フラグメントが他のフラグメントより後で到着しても、前記第１の発明による通信宛先管理が可能になる。

【００３７】（４）本発明の第４は、請求項４および請求項９にそれぞれ記載の通信宛先管理方法および装置であり、前記第１から第３までの発明のゲートウェイに関するものである。図４は第４の発明の原理図である。

【００３８】図４の２８が宛先管理手段２３およびフラグメント処理手段２７とから構成されるデータリンクブリッジ手段である。図４のその他の符号は図１と図３と同じである。

【００３９】第４の発明の方法は、ゲートウェイ２０において下位プロトコル４の内のデータリンク層に、外部ネットワークデータリンク処理部と、内部ネットワークデータリンク処理部と、データリンクブリッジとを設け、データリンクブリッジにより、外部ネットワーク１からのパケットの宛先アドレスを書き換えて、内部ネットワーク２に転送することを特徴とする。

【００４０】第４の発明の装置は、前記第４の発明の方法に記載のゲートウェイ２０であって、下位プロトコル４の内のデータリンク層に、外部ネットワークデータリンク処理手段２１と内部ネットワークデータリンク処理手段２４とデータリンクブリッジ手段２８とを備え、前記第４の発明の方法を実現する。

【００４１】第４の発明によって外部ネットワーク１から内部ネットワーク２に転送されるパケットは、下位プロトコル４であるデータリンク層で宛先アドレスが書き換えられることになり、上位プロトコル３の処理を受けることがなくデータリンク層でデータリンクブリッジを形成することになる。

【００４２】これにより下位プロトコル４だけで外部ネットワーク１および内部ネットワーク２のプロセス間通信が可能になり、ゲートウェイ２０における負荷を軽減することができる。

【００４３】（５）本発明の第５は、請求項５および請求項１０にそれぞれ記載の通信宛先管理方法および装置であり、前記第４の発明のゲートウェイを拡張したものである。図５は第５の発明の原理図である。

【００４４】図５では複数の内部ネットワークデータリンク処理部および複数の内部ネットワークデータリンク処理手段２４が設けられ、複数の内部ネットワーク２がゲートウェイであり本発明の通信宛先管理装置２０に接続できる。図５のその他の符号は図１、図３、図４と同じである。

【００４５】第５の発明の方法は、ゲートウェイにおいて、下位プロトコル４の内のデータリンク層に複数の内部データリンク処理部を設け、複数の内部ネットワークを接続し、複数クラスタ内のプロセス間通信および複数クラスタと外部ネットワークとのプロセス間通信を行うことを特徴とする。

【００４６】第５の発明の装置は前記第５の発明の方法に記載のゲートウェイであって、前記データリンク層に複数の前記内部ネットワークデータリンク処理手段を備え、前記第５の発明の方法を実現する。

【００４７】第５の発明によって該ゲートウェイを一つのハブとしてシステムを構成することが可能になり、該ゲートウェイに一つのポートを割り付けるだけで複数のクラスタ間および外部ネットワークとのプロセス間の通信が可能になる。

【００４８】

【発明の実施の形態】本発明の実施の形態について図面を用いて詳細に説明する。図６はプロセス間通信にＴＣＰ／ＩＰを使用した場合のデータフォーマットおよび本発明の実施の形態であるシステム構成図である。

【００４９】ＴＣＰ／ＩＰでプロセス間通信を行う場合、データはＴＣＰのポートを宛先にして相手先プロセスに送られる。ユーザデータはトランスポート層でＴＣＰヘッダを、ネットワーク層でＩＰヘッダを、データリンク層で物理ヘッダをそれぞれ付与されて物理フレームを形成する。物理フレームでは物理ヘッダとＩＰヘッダの長さが決まっているので物理フレーム内のＴＣＰの宛先ポートの位置を特定することができる。

【００５０】図６の１は外部のネットワーク、２は内部のネットワークであり、それぞれローカルエリアネットワーク（ＬＡＮ）を構成する。図６の１１はワークステーションであり図１のホストコンピュータもしくはプロセッサエレメント１０に対応する。

【００５１】図６の６０はゲートウェイを兼ねるプロセッサエレメントであり、図１のゲートウェイである本発明通信宛先管理装置２０に対応する。３１はプロセッサエレメントであり、図１のホストコンピュータもしくはプロセッサエレメント３０に対応する。４０はゲートウェイ６０および複数のプロセッサエレメント３１で構成される並列システムである。

【００５２】図６の６１と６４はそれぞれ外部ＬＡＮデータリンクと内部ＬＡＮデータリンクであり、図１の外部ネットワークデータリンク処理手段２１と内部ネットワークデータリンク処理手段２４に対応する。６２はポート・ＰＥ対応表であり、図１のポート・アドレス対応表６に対応する。

【００５３】図６の６８はデータリンクブリッジであり、宛先管理手段２３とフラグメント処理手段２７から構成される。６９はプロセッサエレメント３１にあるポ

ート管理クライアントであり、ゲートウェイ６０にあるポート管理サーバ２５と共に並列システム内でのプロセスで使用されるポートを管理する。

【００５４】図６において、ワークステーション１１で実行されるプロセスＡから並列システム４０内のプロセッサエレメント３１で実行されるプロセスＢへ、ＴＣＰ／ＩＰで通信を行う場合、プロセスＡは、プロセスＢのトランスポート層のポートを宛先に通信を行う。

【００５５】送信データは、ワークステーション１１においてプロセスＡからトランスポート層プロトコル処理、ネットワーク層プロトコル処理、およびデータリンク処理を通して、外部ＬＡＮに送出される。

【００５６】並列システム４０のゲートウェイ６０では、外部ＬＡＮデータリンク６１においてこの送信データを受取り外部ＬＡＮ用のデータリンク処理を行い、データリンクブリッジ６８へこの物理フレームを渡す。

【００５７】データリンクブリッジ６８内の宛先管理手段２３ではトランスポート層やネットワーク層のそれぞれのプロトコル処理を行わずに、物理フレーム上の宛先ポートの位置を直接読み、ポートがどのプロセッサエレメントに対応付けられているかをポート・ＰＥ対応表６２から判断し、物理フレームにある物理ヘッダの宛先アドレスを書き換えてから、その物理フレームを内部ＬＡＮデータリンク６４に渡す。

【００５８】内部ＬＡＮデータリンク６４では、内部ＬＡＮ用のデータリンク処理を行い、内部ＬＡＮにパケットを送出する。プロセスＢを実行するプロセッサエレメント３１ではパケットを受け取り、データリンク処理、ネットワーク層プロトコル処理、およびトランスポート層プロトコル処理を行い、プロセスＢに渡す。

【００５９】ポート・アドレス対応表６２は、並列システム４０の中のポート管理クライアント６９およびポート管理サーバ２５によって管理される。図７にポート管理クライアントサーバの詳細を示す。

【００６０】ポート管理クライアント６９はプロセッサエレメント３１の中にあり、プロセスとのインタフェースを持つインタフェース部と、ポート管理サーバ２５の通信部と通信を行う通信部とから成る。

【００６１】インタフェース部は既存のポート関連の機能インタフェースを通して、プロセスからのポートの割り付け要求や削除要求を受け付け、ポートの割り付けや削除を行う。ポート管理クライアント６９の通信部は、ポート割り付け要求や削除要求をポート管理サーバ２５に送ったり、ポート管理サーバからの新規番号の通知を受け取ったりする。

【００６２】ポート管理サーバ２５はゲートウェイ６０の中にあり、ポート管理クライアント６９の通信部と通信を行う通信部と、ポート・ＰＥ対応表６２の検索、登録および削除を行う対応部とから成る。

【００６３】ポート管理サーバ２５の通信部は、ポート管理クライアントからのポート割り付け要求や削除要求を受け取ったり、ポート管理サーバへ新規番号を通知したりする。対応部は、ポート管理クライアントからの要求に従い、ポート・ＰＥ対応表６２のエントリを検索したり、登録したり、削除したりする。

【００６４】ポート管理クライアントサーバは以下の手順で実行される。まず並列システム４０内のプロセス（図７のプロセスＢ）がポートの割り付け要求を発行すると、ポート管理クライアントが、ゲートウェイ６０のポート管理サーバ２５にポートの発行および登録を依頼する。

【００６５】依頼を受けたポート管理サーバ２５は、ポート・ＰＥ対応表６２を参照し、未使用のポートを選び、その対応関係をポート・ＰＥ対応表６２に登録すると共に選ばれたポートをポート管理クライアント６９に通知する。ポート管理クライアント６９はプロセスＢにそのポートを割り付ける。

【００６６】ポートを開放する場合や、プロセスの終了時には、ポート・ＰＥ対応表６２のエントリはポート管理サーバ２５によって削除される。またポート管理サーバ２５はプロセスＢの要求に応じて対応表を検索することもある。

【００６７】ＵＮＩＸアプリケーションではｓｏｃｋｅｔ等により、トランスポート層プロトコルであるｂｉｎｄ、ｃｏｎｎｅｃｔなどの既存のインタフェースを使ってポートを割り付ける。

【００６８】図８はフラグメント宛先管理構成図である。図８の（Ａ）はＴＣＰ／ＩＰプロトコルにおけるフラグメントの例であり、図３の第３の発明の原理図（Ａ）に対応する。ＩはＴＣＰ／ＩＰのパケット識別子であるＩＰ－ＩＤＥＮＴＩＦＩＣＡＴＩＯＮ（ＩＰ－ＩＤ）であり、Ｆは同じく順番識別子であるＦＬＡＧＭＥＮＴＯＦＦＳＥＴ（オフセット）である。ＩとＰは共に発信元のネットワーク層（ＩＰ層）で付与される。Ｐはトランスポート層（ＴＣＰ層）の宛先ポートである。

【００６９】物理フレームの（ａ）が先頭フラグメントであり、（ｂ）が先頭以外のフラグメントである。先頭フラグメント（ａ）には宛先ポートが含まれる。物理フレームの転送の場合は、フラグメントの順序性が保証されていないハードウェアプロトコルがあり、（ｂ）が先に到着する場合がある。

【００７０】そのためにフラグメント化する時に全フラグメントには、パケット固有の識別子であるＩＰ－ＩＤと、フラグメントの順番を識別するためのオフセットが付与される。オフセットによりフラグメントの順番を知ることができる。

【００７１】図８の（Ｂ）はフラグメント宛先管理構成図である。６６はフラグメントバッファであり図３のフラグメント記憶手段２６に対応する。２７は図３乃至図６に示されるフラグメント処理手段２７と同じである。

フラグメント処理手段２７は、データリンク層プロトコルで外部ＬＡＮデータリンク６１および内部ＬＡＮデータリンク６４に接続される。

【００７２】またフラグメント処理手段２７は、ポート・ＰＥ対応表６２に接続され、先頭フラグメントに含まれる宛先ポートに対応するプロセッサアドレス（ＰＥアドレス）を読み出して、物理ヘッダの宛先アドレスを書き換える。

【００７３】図８の６７はＩＰ－ＩＤ・ＰＥ対応表である。ＩＰ－ＩＤ・ＰＥ対応表６７はフラグメント処理手段２７が、フラグメントの物理ヘッダの宛先アドレスを書き換えるために、ＩＰ－ＩＤとＰＥアドレスの対応を記憶しておくためにフラグメント処理手段２７の中に備えられる。

【００７４】図９はフラグメント処理手段２７におけるフラグメント処理フロー図である。ステップｓ１では、オフセットにより先頭フラグメントかどうかを判定する。先頭フラグメントであればステップｓ２に進み、先頭フラグメントでなければステップｓ３に進む。

【００７５】ステップｓ２では、先頭フラグメントに含まれる宛先ポートにより、対応するＰＥアドレスをポート・ＰＥ対応表６２から読み出す。続いてＩＰ－ＩＤと読み出したＰＥアドレスの対応をＩＰ－ＩＤ・ＰＥ対応表６７に登録し、ステップｓ５に進む。

【００７６】ステップｓ３では、ＩＰ－ＩＤ・ＰＥ対応表６７を参照しＩＰ－ＩＤと宛先ＰＥが既に登録済みかどうかを判定する。ＩＰ－ＩＤと宛先ＰＥが既に登録済みであれば、ステップｓ５に進み、未だ登録してなければステップｓ４に進む。

【００７７】ステップｓ４では、そのフラグメントをフラグメントバッファ６６に格納し、先頭フラグメント待ちのためにステップｓ１に戻る。ステップｓ５では、ＩＰ－ＩＤ・ＰＥ対応表６７を読出し、各フラグメントの物理フレームの物理ヘッダ部分にある宛先のプロセッサエレメントのＰＥアドレスを書き換え、ステップｓ６に進む。

【００７８】ステップｓ６では、フラグメントバッファ６６に先頭待ちの同じＩＰ－ＩＤを持つフラグメントがあるかどうかを判定する。同一ＩＰ－ＩＤを持つフラグメントがあれば、ステップｓ７に進み、同一ＩＰ－ＩＤを持つフラグメントがなければフラグメント処理を終了する。

【００７９】ステップｓ７では、フラグメントバッファ６６から処理待ちのフラグメントを取り出しステップｓ５に戻る。図１０にデータリンクブリッジの構成図を示す。図１０の（Ａ）は、通信宛先管理がデータリンク層で行われることを示している。図の矢印（１）は、外部ＬＡＮからのパケットが通信宛先管理により物理アドレスを書き換えられてトランスポート層やネットワーク層を経由することなしに内部ＬＡＮに送出され、データリンク層でブリッジしていることを示している。

【００８０】矢印（２）は、外部ＬＡＮからのパケットが通信宛先管理により、ゲートウェイを兼ねたプロセッサエレメント６０に宛てたものと判定され、上位層であるネットワーク層やトランスポート層に渡されることを示している。

【００８１】矢印（３）は、内部ＬＡＮ２からのパケットが外部ＬＡＮに接続されるワークステーション１１や他のホストコンピュータやプロセッサエレメントが宛先であるので通信宛先管理を経由せず、外部ＬＡＮに渡されることを示している。

【００８２】矢印（４）も、パケットの宛先が外部ＬＡＮに接続されるワークステーション１１や他のホストコンピュータやプロセッサエレメントパケットであるので、プロセッサエレメント６０から通信宛先管理を経由せず、外部ＬＡＮに渡されることを示している。

【００８３】矢印（５）と（６）はそれぞれ内部ＬＡＮからプロセッサエレメント６０へのパケットおよびプロセッサエレメント６０から内部ＬＡＮへのパケットであり、やはり通信宛先管理を経由しないで、すなわちデータリンクブリッジを通さないで、それぞれプロセッサエレメント６０の上位層および内部ＬＡＮに渡されることを示している。

【００８４】図１０の（Ｂ）は、通信宛先管理手段２３およびフラグメント処理手段２７から構成されるデータリンクブリッジ６８が、データリンク層で外部ＬＡＮデータリンク６１および内部ＬＡＮデータリンク６４に接続され、またポート・ＰＥ対応表６２を参照して通信の宛先管理をしていることを示している。

【００８５】図１１はデータリンクブリッジの処理フロー図である。まずステップｓ１０で外部ＬＡＮデータリンクからパケットを受け取りステップｓ１１に進む。この時のパケットのフォーマットは物理フレームである。

【００８６】ステップｓ１１では、その物理フレームがフラグメント化しているかどうかを判定する。フラグメント化していなければ、ステップｓ１２に進み、フラグメント化している場合はステップｓ１３に進む。

【００８７】ステップｓ１２では、物理フレームの宛先ポート位置を算定し、その宛先ポートに基づいてポート・ＰＥ対応表６２を読出し、宛先の物理アドレスを書き換える。次にステップｓ１４に進む。

【００８８】ステップｓ１３では、図９のフラグメントの処理フロー図に示す処理を行い、宛先の物理アドレスを書き換える。次にステップｓ１４に進む。ステップｓ１４では、自ノードに宛てたものかどうかを判定する。宛先が自ノードでなければステップｓ１５へ進み、宛先が自ノードであればステップｓ１６へ進む。

【００８９】ステップｓ１５では、内部ＬＡＮにパケットを送出すべく内部ＬＡＮデータリンクへ物理アドレスを書き換えた物理フレームを渡して処理を終了する。ス

15

テップｓ１６では、データリンク処理を行い、自ノードすなわちプロセッサエレメント６０の上位層にデータを渡して処理を終了する。

【００９０】図１２は複数の内部ＬＡＮデータリンク接続図である。図１２では複数の内部ＬＡＮデータリンク６４がデータリンク層に設けられ、複数の内部ＬＡＮ２がゲートウェイであるプロセッサエレメント６０に接続できることを示している。

【００９１】図１３は複数のクラスタの通信宛先管理システム構成図である。並列システム４０が複数の独立した内部ネットワーク（クラスタ）から構成され、プロセッサエレメント６０をゲートウェイとして外部ＬＡＮ１を介してワークステーション（ＷＳ）に接続されている。

【００９２】プロセッサエレメント６０にはポート・ＰＥ対応表６２を備えており、内部のプロセッサエレメントで実行されるプロセスに対応する、プロセッサの物理アドレスを管理している。従って外部ＬＡＮに接続された例えばワークステーションから見れば、並列システムの中のプロセッサエレメントの物理アドレスを宛先としないでもよいことになる。

【００９３】すなわち並列システム４０の中のゲートウェイであるプロセッサエレメント６０の物理アドレスを宛先にすれば、並列システム４０の中でプロセスがどのプロセッサで実行されているかを問題にしなくて済むことになり、外部ＬＡＮからのプロセス間通信では並列システム４０の中のプロセッサのＩＰアドレスは一つでよいことになる。

【００９４】同様に並列システム４０の内部でのクラスタ間のプロセス間通信でも、プロセッサのＩＰアドレスは一つでよいことになる。また同様に並列システム４０の内部でのクラスタ内のプロセス間通信でも、プロセッサのＩＰアドレスは一つでよいことになる。

【００９５】図１４に内部ネットワークに接続される並列システムの構成図を示す。図１４では、本発明の通信宛先管理装置をホストコンピュータに適用し、プロセスを行うプロセッサエレメント（ＰＥ）を内部ネットワークを介して接続している。図１４の６２と６４は、それぞれポート・ＰＥ対応表と内部ＬＡＮデータリンクである。

【００９６】尚、本発明の通信宛先管理装置は一つのホストコンピュータあるいは一つのプロセッサエレメントを兼ねたゲートウェイであってもよいし、ゲートウェイ専用装置であってもよいことは言うまでもない。図１５に複数クラスタを接続するハブ（ＨＵＢ）を用いた通信宛先管理システムの構成図を示す。

【００９７】また外部ネットワーク１に他の並列システムあるいは分散システムを接続して大規模な並列システムあるいは分散システムを構成することができる。それぞれの並列システムもしくは分散システムでは、内部で

16

ネットワークを構成しており、外部ネットワークを介してそれぞれのゲートウェイで接続される。図１６に大規模分散処理における通信宛先管理システム構成図を示す。

【００９８】図１６ではプロセスＡからプロセスＢへの通信の経路を示した。また図１６のように発信元（プロセスＡ側）の並列システムでも本発明のゲートウェイをインストールすれば、プロセスＢからプロセスＡへの通信にも本発明の通信宛先管理が適用可能になり、大規模で通信効率の良い並列システムもしくは分散システムを構築できる。

【００９９】また図１６のように外部ネットワーク接続されるゲートウェイに本発明の通信宛先管理の方法および装置を適用し、内部ネットワークで接続される並列システム内では、それぞれのシステムで独自のプロトコルに基づくプロセス間通信を行うこともできる。

【０１００】外部ネットワーク上ではＴＣＰ／ＩＰなどの一般プロトコルに基づく通信を行い、それぞれの内部ネットワーク上のプロセス間通信では、内部で固有の宛先プロセスのアドレスを用いる。これにより異機種間でのプロセス間通信を外部ネットワークを通じて行うことが可能になる。

【０１０１】

【発明の効果】以上の説明から明らかなように本発明によれば、下位プロトコルに属するデータリンク層で通信の宛先管理を実行することができるので、上位プロトコルに必須の処理がゲートウェイでは不要になり、ゲートウェイの負荷を低減し通信の相手先のプロセッサに負荷を分散させることが可能になる、内部ネットワークに１つのＩＰアドレスを割り当てることができるのでＩＰアドレスを増やすことなしにホストコンピュータもしくはプロセッサエレメントを増やすことが可能になると共に、並列システムもしくは分散システムを１つのシステムとして扱うことが可能になる、プロセス間通信で扱う宛先アドレスを常に自動的に更新および検索できるのでユーザは今まで通りの使用法でプロセス間通信を行える、フラグメント化に対処しているのでフラグメントが発生してもフラグメントの順序性を確保できる、ゲートウェイのみならず従来からのハブにも適用することができるので、複数クラスタの通信宛先管理が可能になり、更に大規模な並列もしくは分散システムを構築できるという様々な効果がある。

【図面の簡単な説明】

【図１】　第１の発明の原理図

【図２】　第２の発明の原理図

【図３】　第３の発明の原理図

【図４】　第４の発明の原理図

【図５】　第５の発明の原理図

【図６】　本発明実施の形態のＴＣＰ／ＩＰデータフォーマットおよびシステム構成図

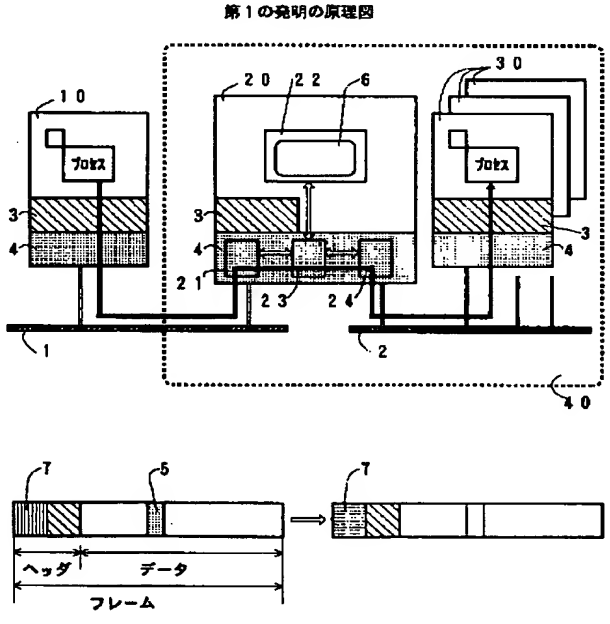
17

【図７】　ポート管理クライアントサーバ
【図８】　フラグメントの宛先管理構成図
【図９】　フラグメント処理フロー図
【図１０】　データリンクブリッジ構成図
【図１１】　データリンクブリッジ処理フロー図
【図１２】　複数の内部ＬＡＮデータリンク接続図
【図１３】　複数クラスタの通信宛先管理システム構成図
【図１４】　内部ネットワークで構成される並列システム構成図
【図１５】　ハブを用いた通信宛先管理システム構成図
【図１６】　独自プロセス間通信による通信宛先管理システム構成図
【図１７】　ＩＳＯの７層参照モデル
【図１８】　ＴＣＰ／ＩＰ　Ｉｎｔｅｒｎｅｔ階層モデル
【図１９】　従来のプロセス間通信の説明図
【符号の説明】
１　外部ネットワーク
２　内部ネットワーク
３　上位プロトコル
４　下位プロトコル
５　宛先ポート情報
６　ポート情報に関するアドレス対応表

18

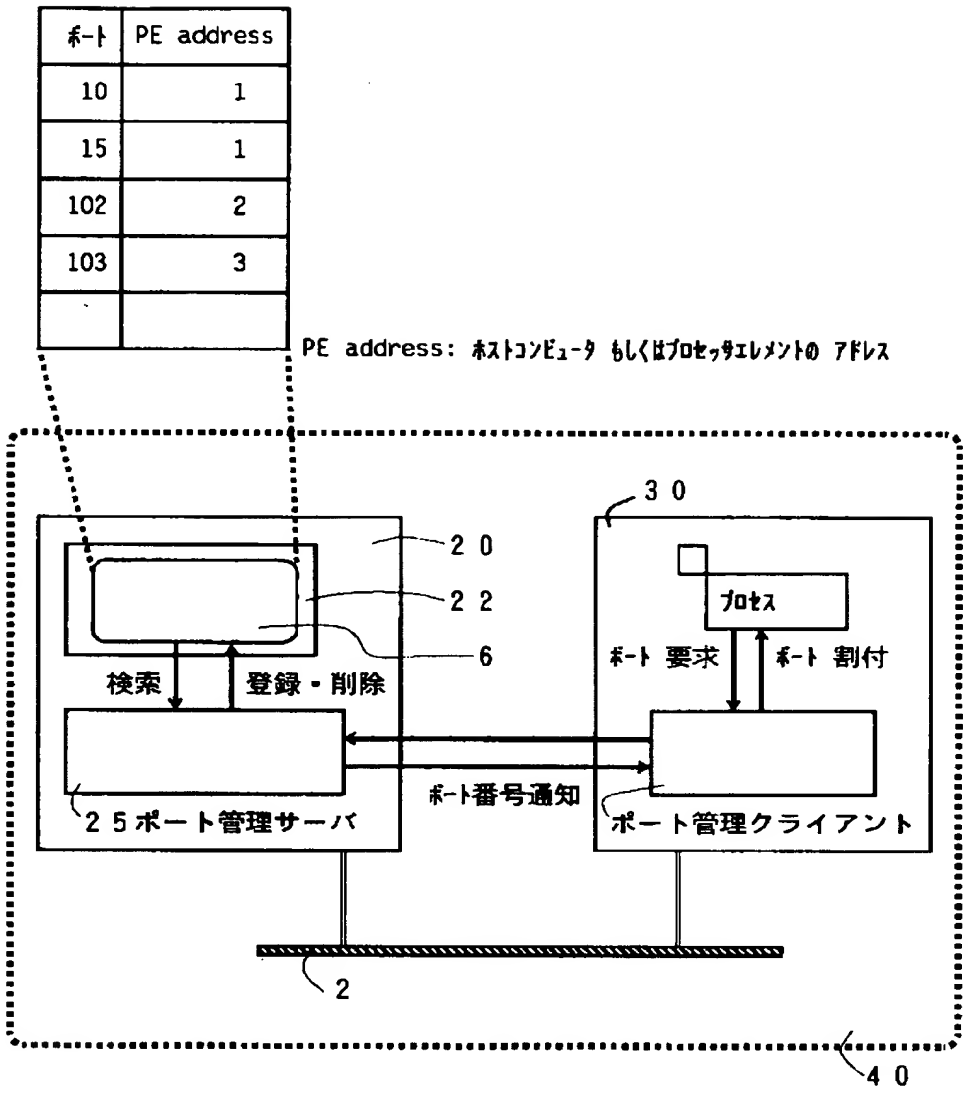
７　宛先アドレス
１０　ホストコンピュータもしくはプロセッサエレメントもしくはゲートウェイ
１１　ワークステーション
２０　ゲートウェイであり、本発明の通信宛先管理装置
２１　外部ネットワークデータリンク処理手段
２２　対応表記憶手段
２３　宛先管理手段
２４　内部ネットワークデータリンク処理手段
２５　ポート管理サーバ
２６　フラグメント記憶手段
２７　フラグメント処理手段
２８　データリンクブリッジ手段
３０　ホストコンピュータもしくはプロセッサエレメント
４０　並列あるいは分散システム
６０　ゲートウェイを兼ねたプロセッサエレメント
６１　外部ＬＡＮデータリンク
６２　ポート・ＰＥ対応表
６４　内部ＬＡＮデータリンク
６６　フラグメントデータバッファ
６７　ＩＰ－ＩＤ・ＰＥ対応表
６８　データリンクブリッジ
６９　ポート管理クライアント

【図１】



【図２】

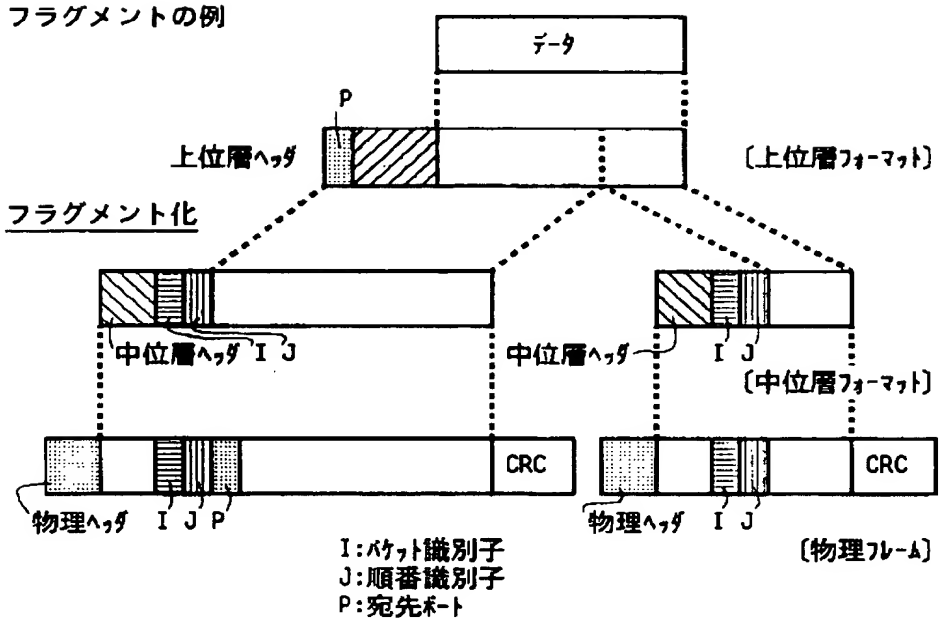
第２の発明の原理図



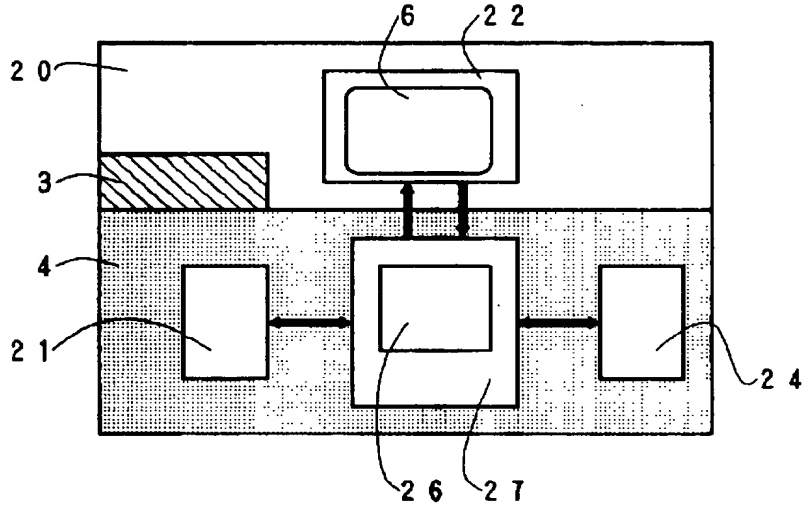
【図３】

第３の発明の原理図

(A) フラグメントの例



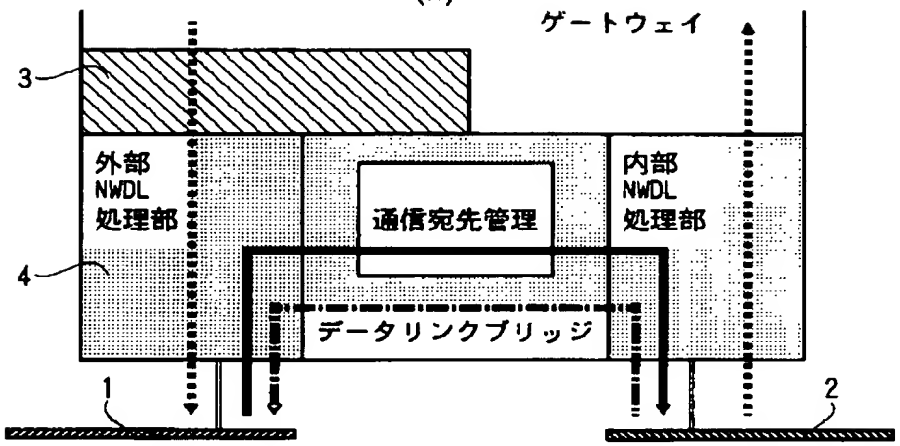
(B) フラグメントの宛先管理



【図４】

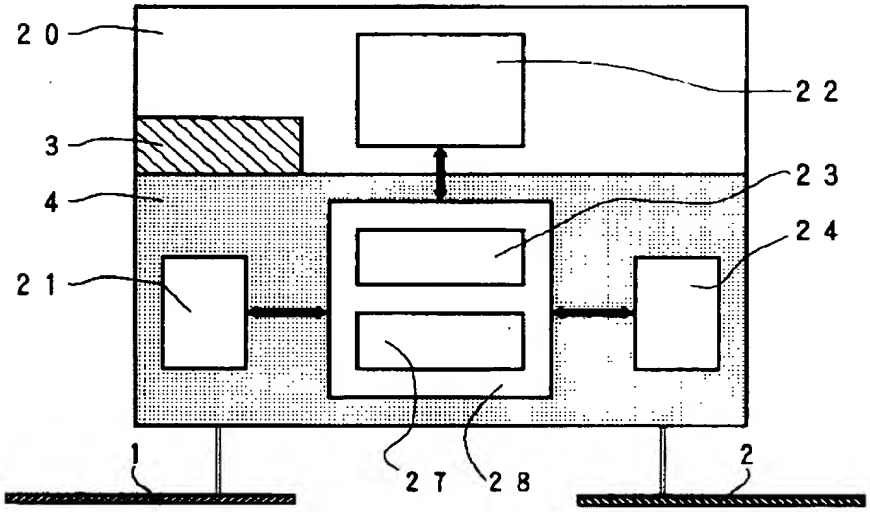
第４の発明の原理図

(A)



NWDL：ネットワークデータリンク

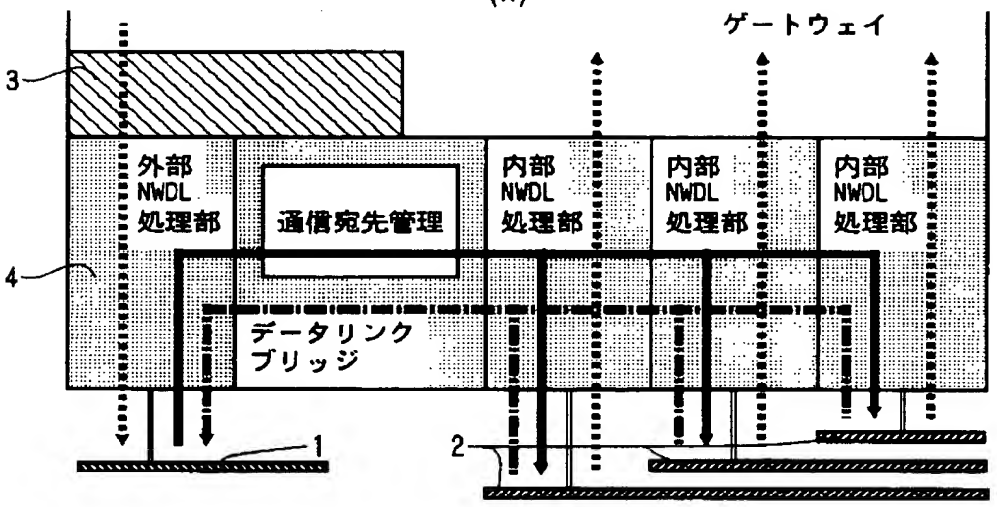
(B)



【図５】

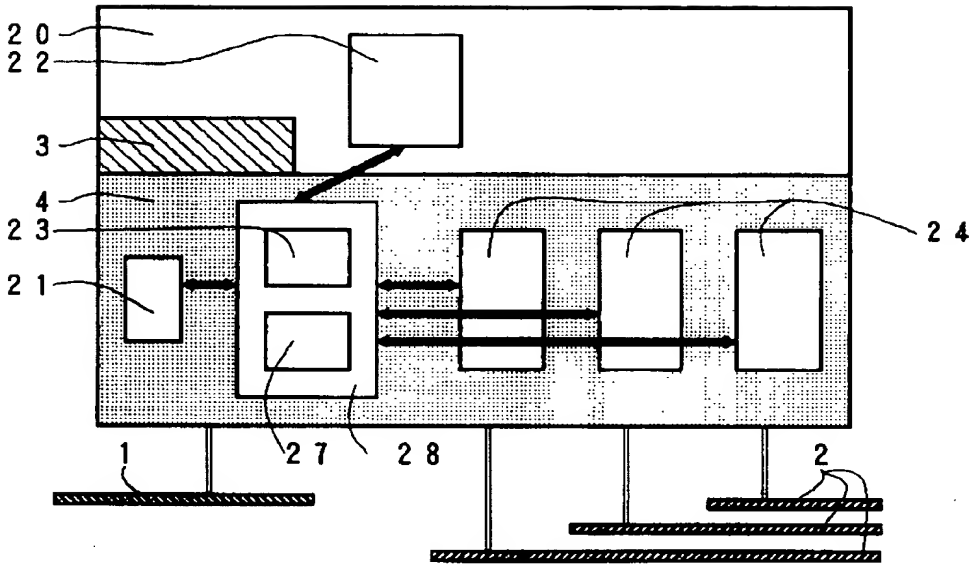
第５の発明の原理図

(A)



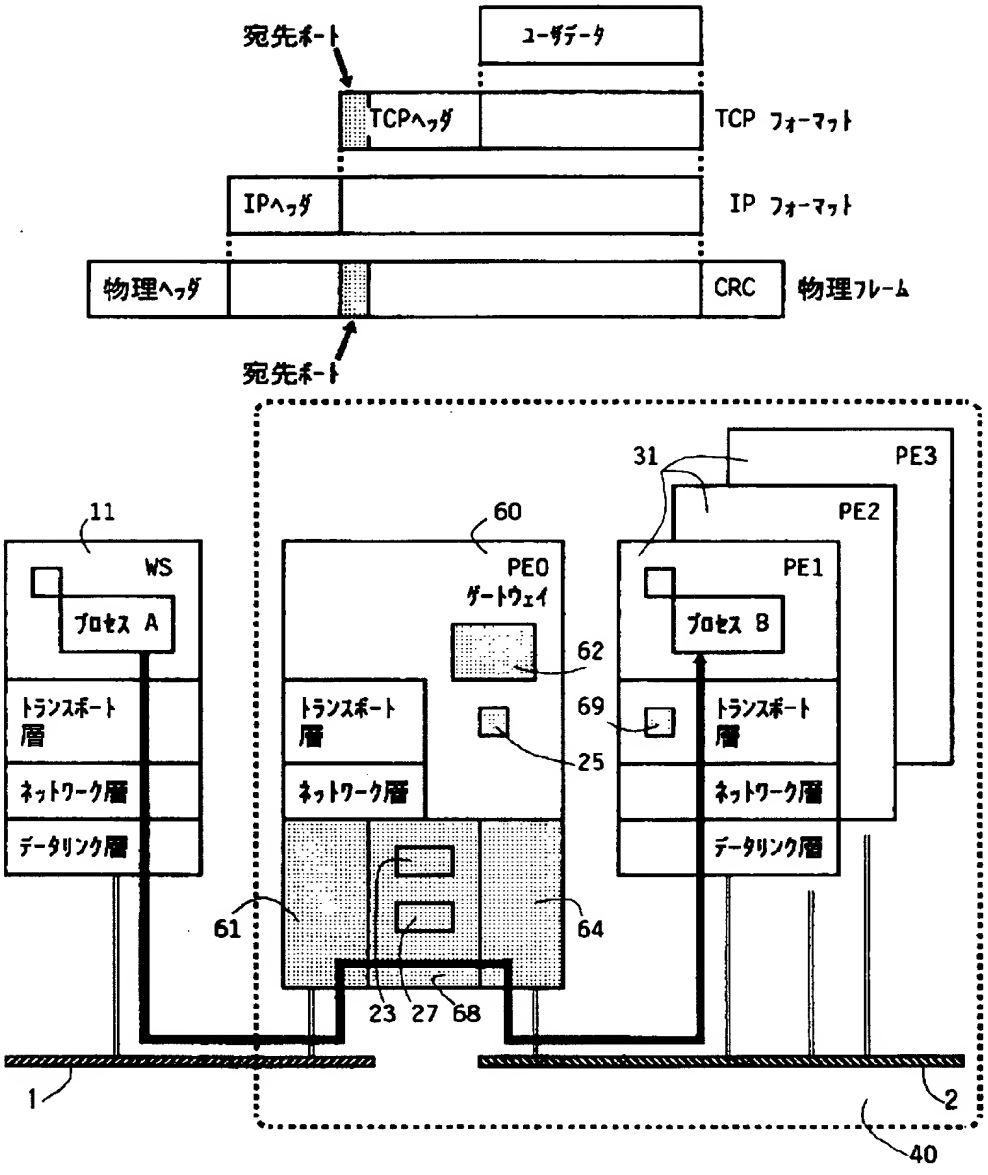
NWDL：ネットワークデータリンク

(B)



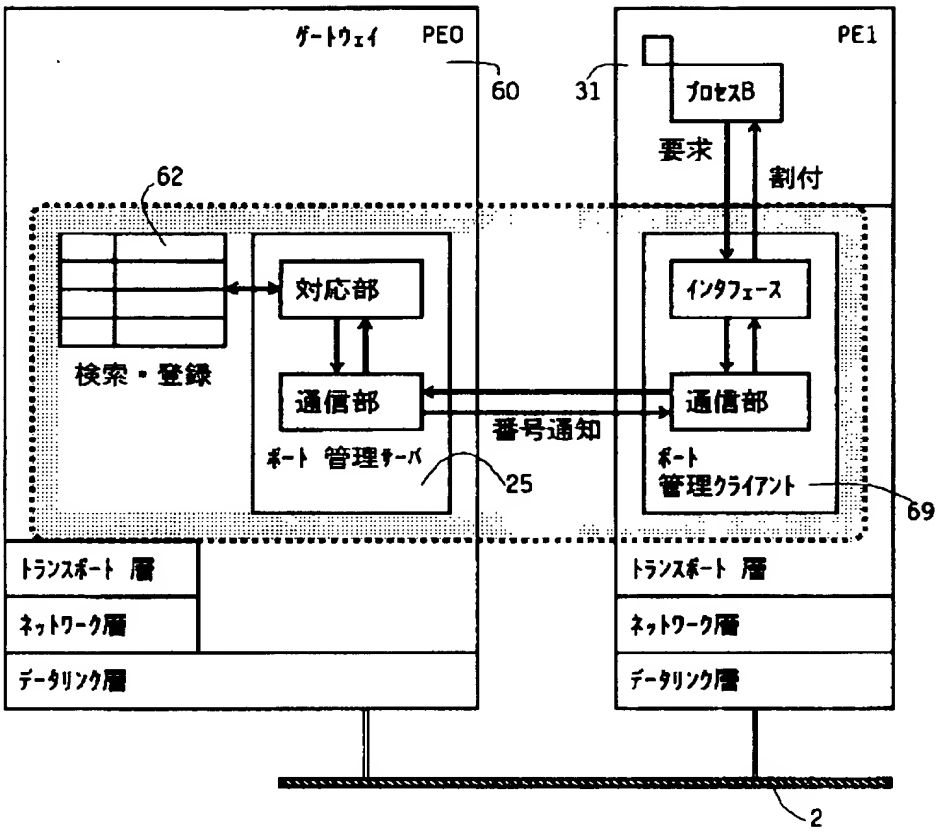
【図６】

ＴＣＰ／ＩＰデータフォーマットおよびシステム構成図



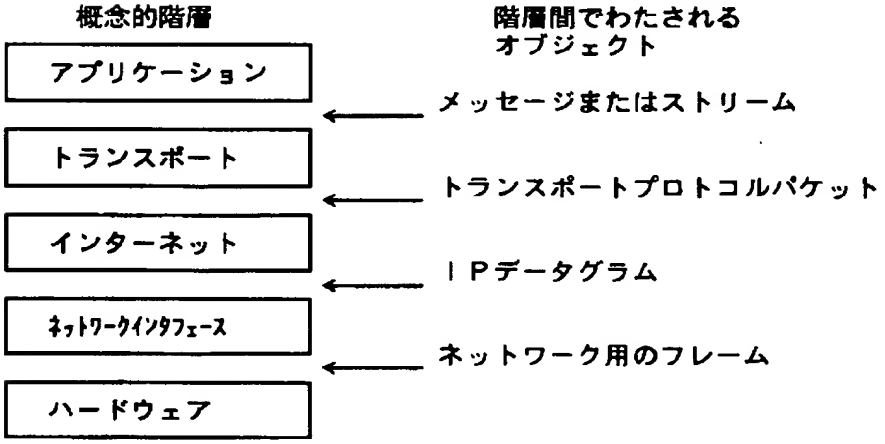
【図７】

ポート管理クライアントサーバ



【図１８】

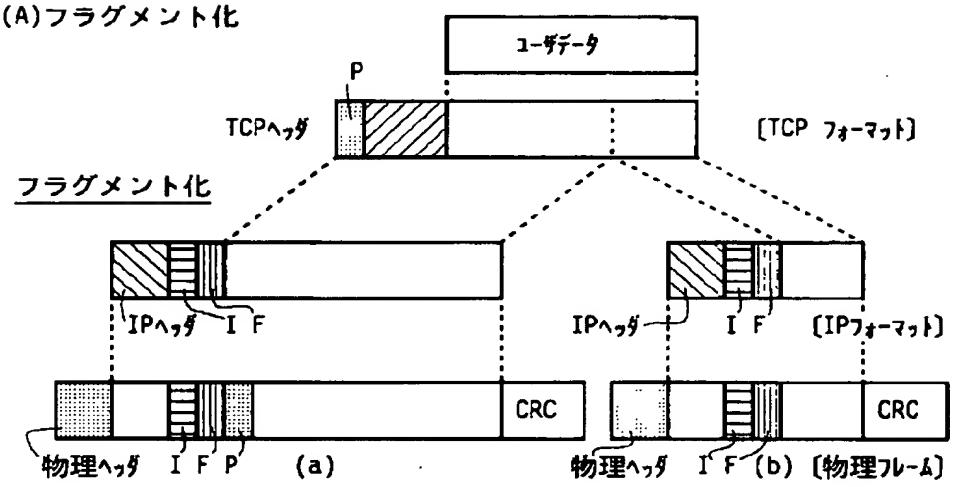
TCP/IP Internet 階層モデル



【図８】

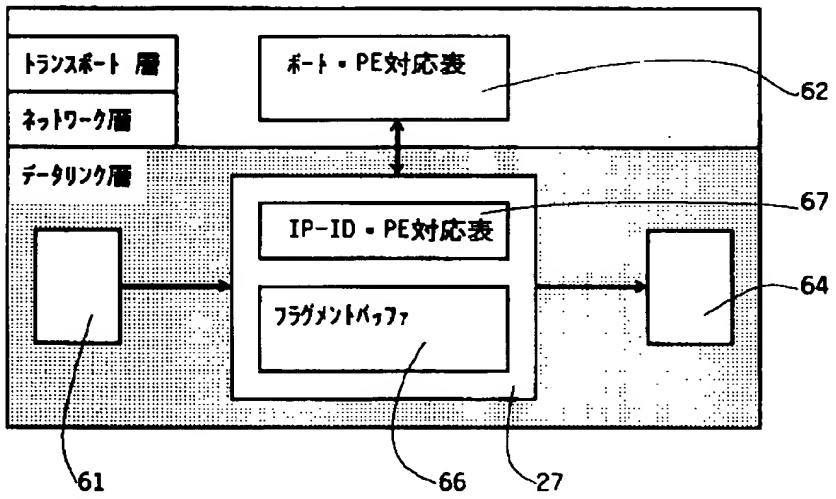
フラグメント宛先管理構成図

(A)フラグメント化



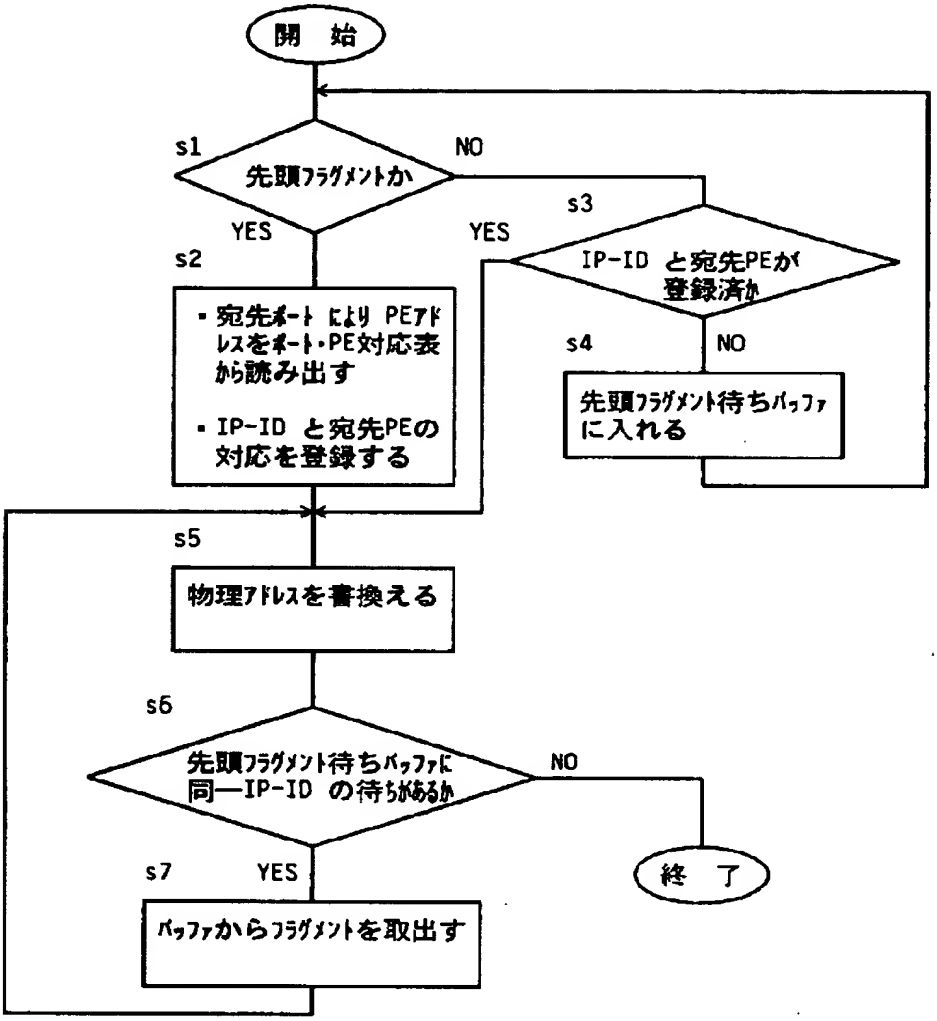
I: TCP/IP のパケット識別子である IP-IDENTIFICATION
F: TCP/IP の順番識別子である FLAGMENT OFFSET
P:　宛先ポート

(B)フラグメント宛先管理構成図



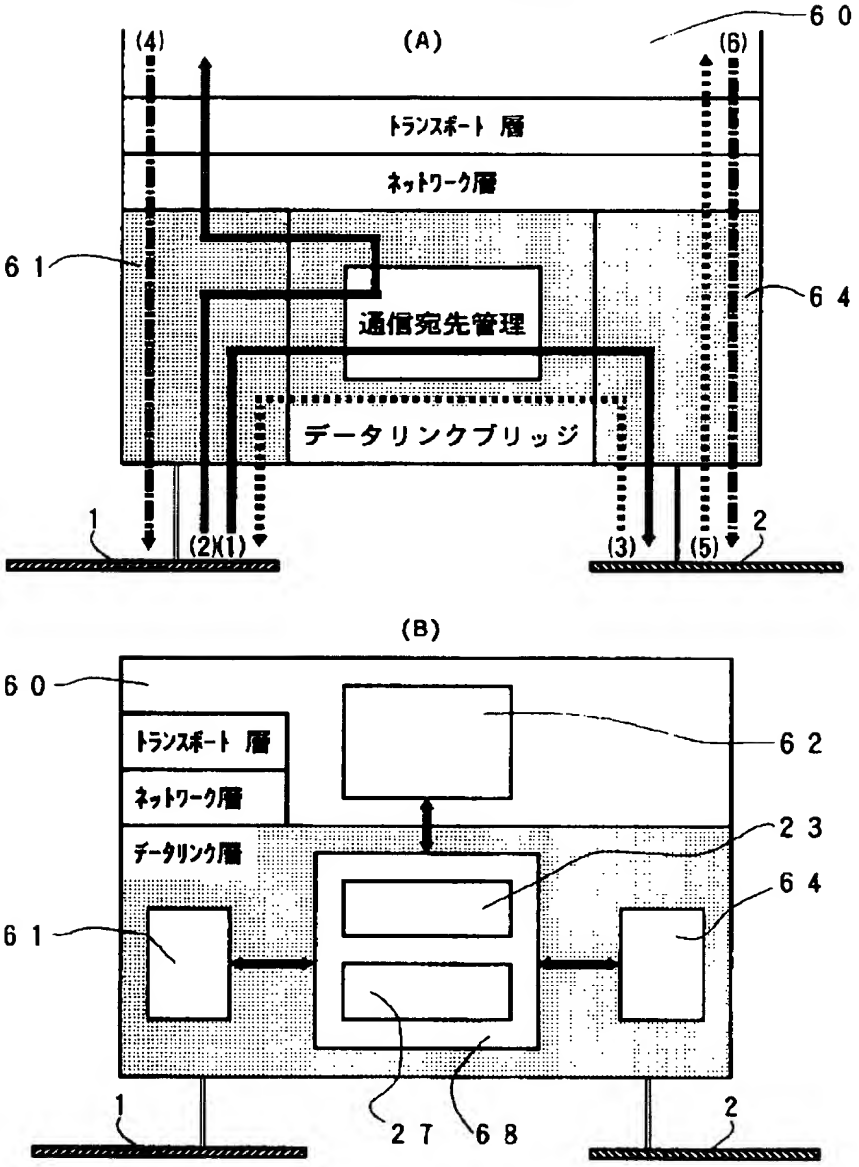
【図９】

フラグメント処理フロー図



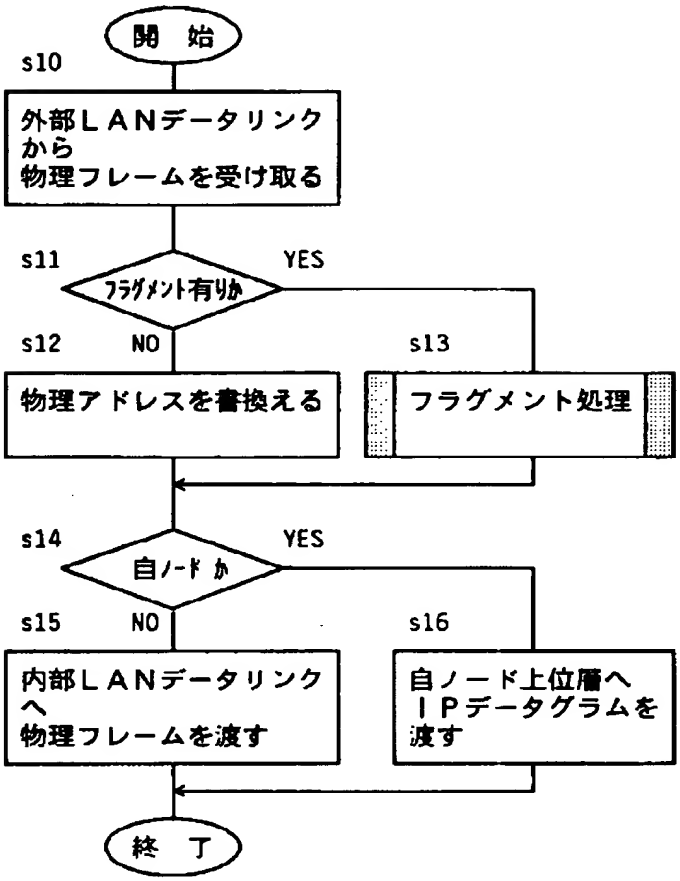
【図１０】

データリンクブリッジ構成図



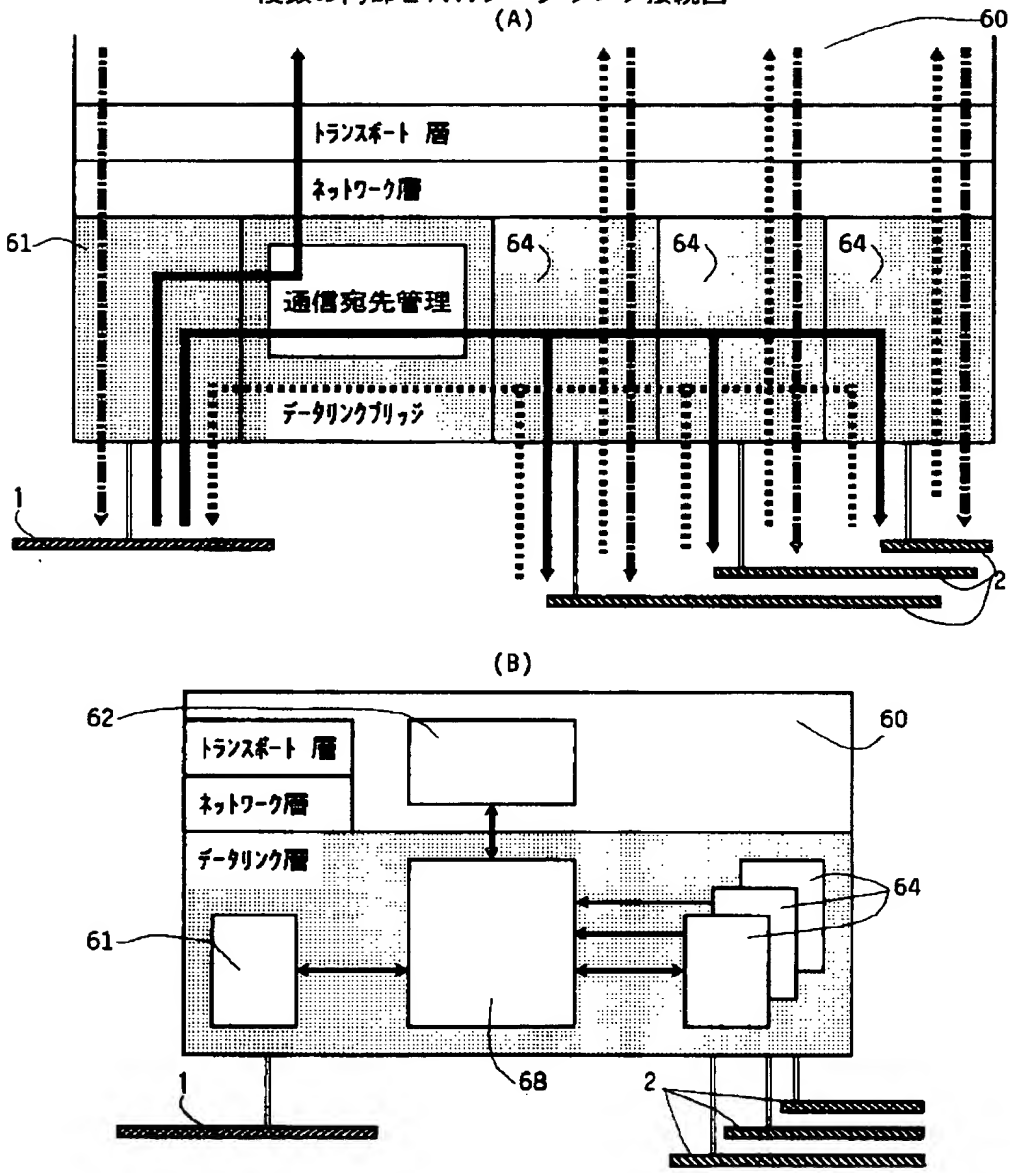
【図１１】

データリンクブリッジ処理フロー図



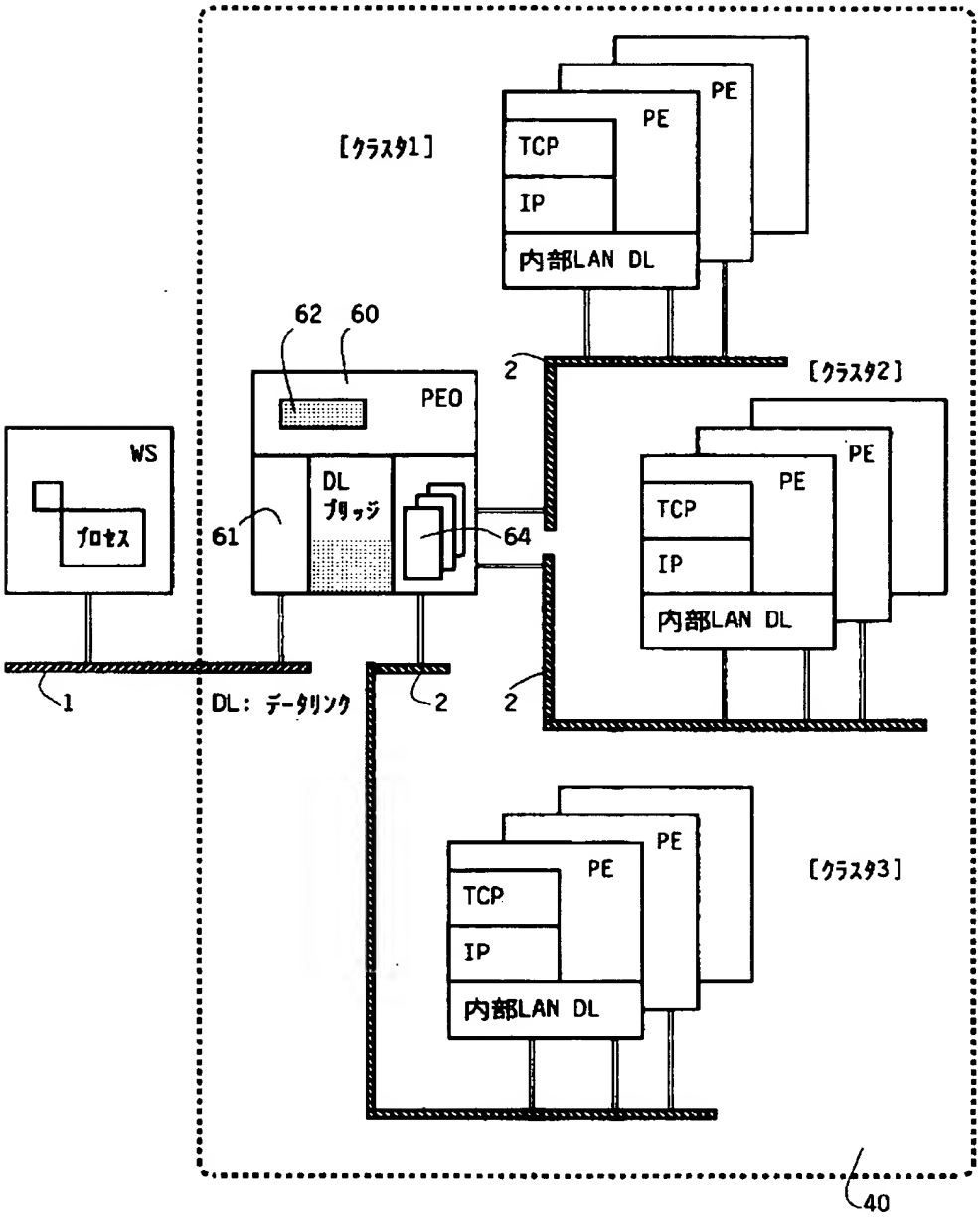
【図１２】

複数の内部ＬＡＮデータリンク接続図



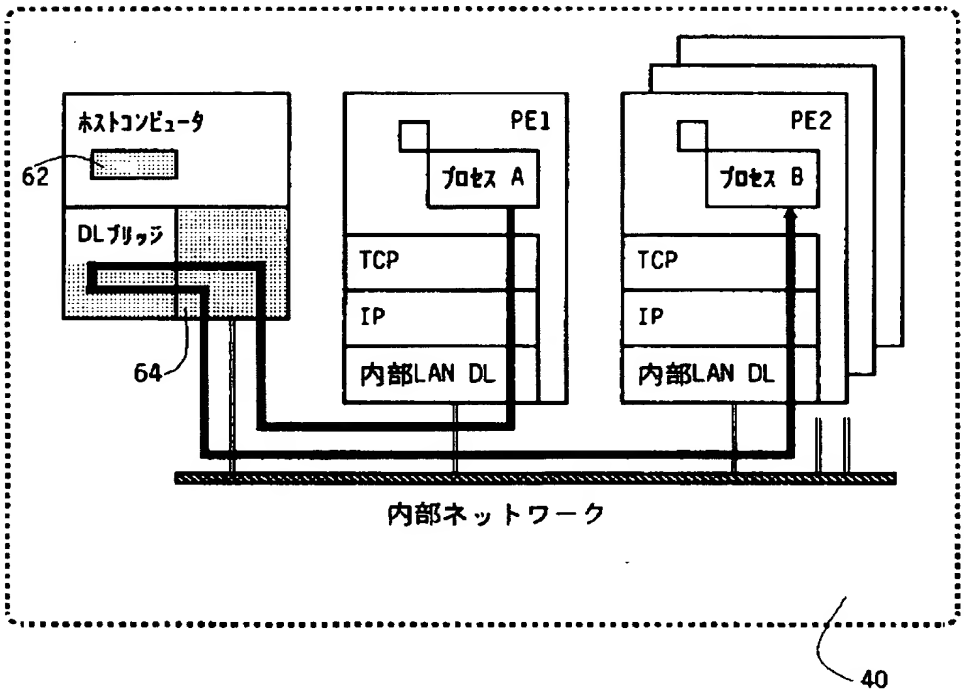
【図１３】

複数クラスタの通信宛先管理システム構成図



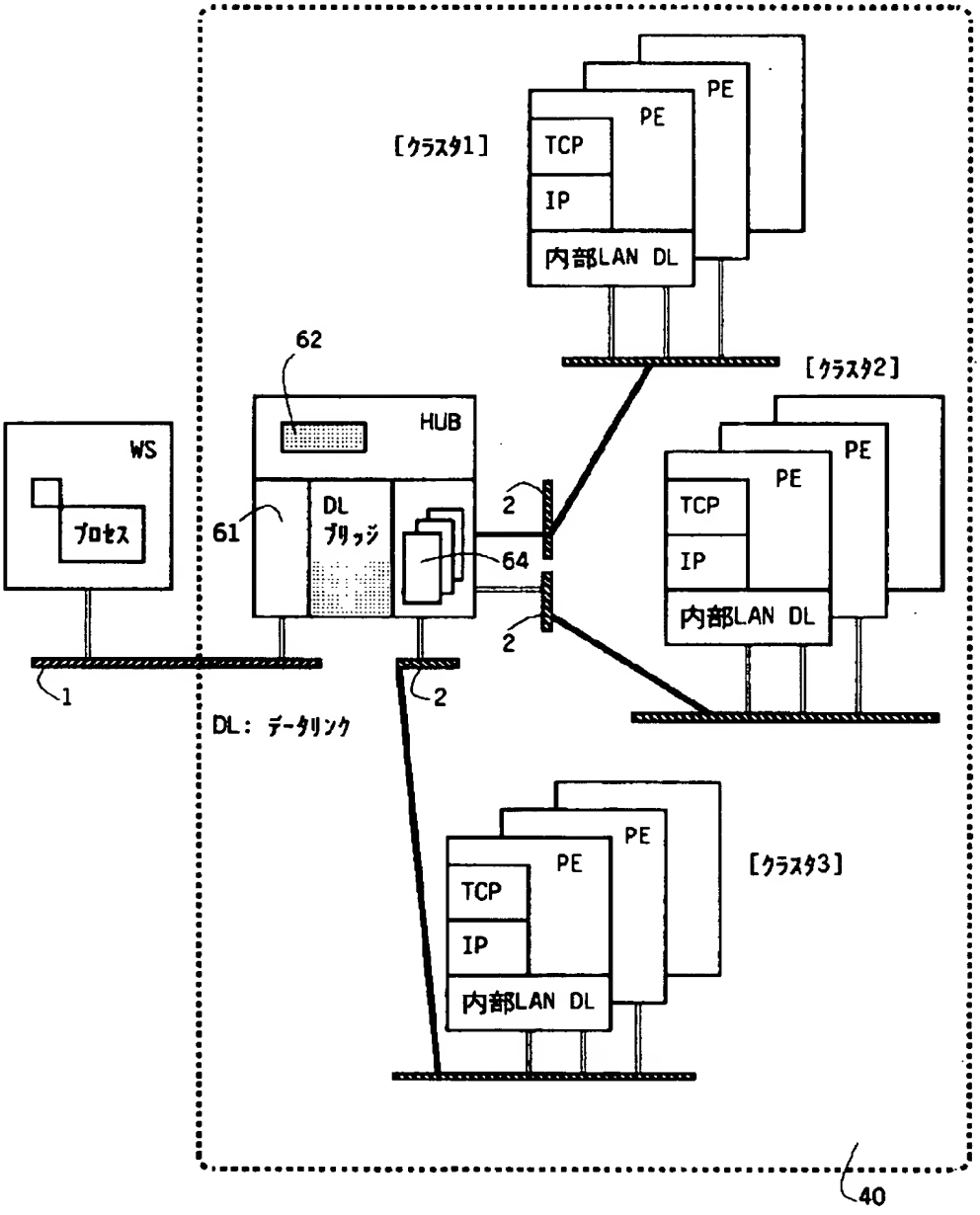
【図１４】

内部ネットワークで構成される並列システム構成図



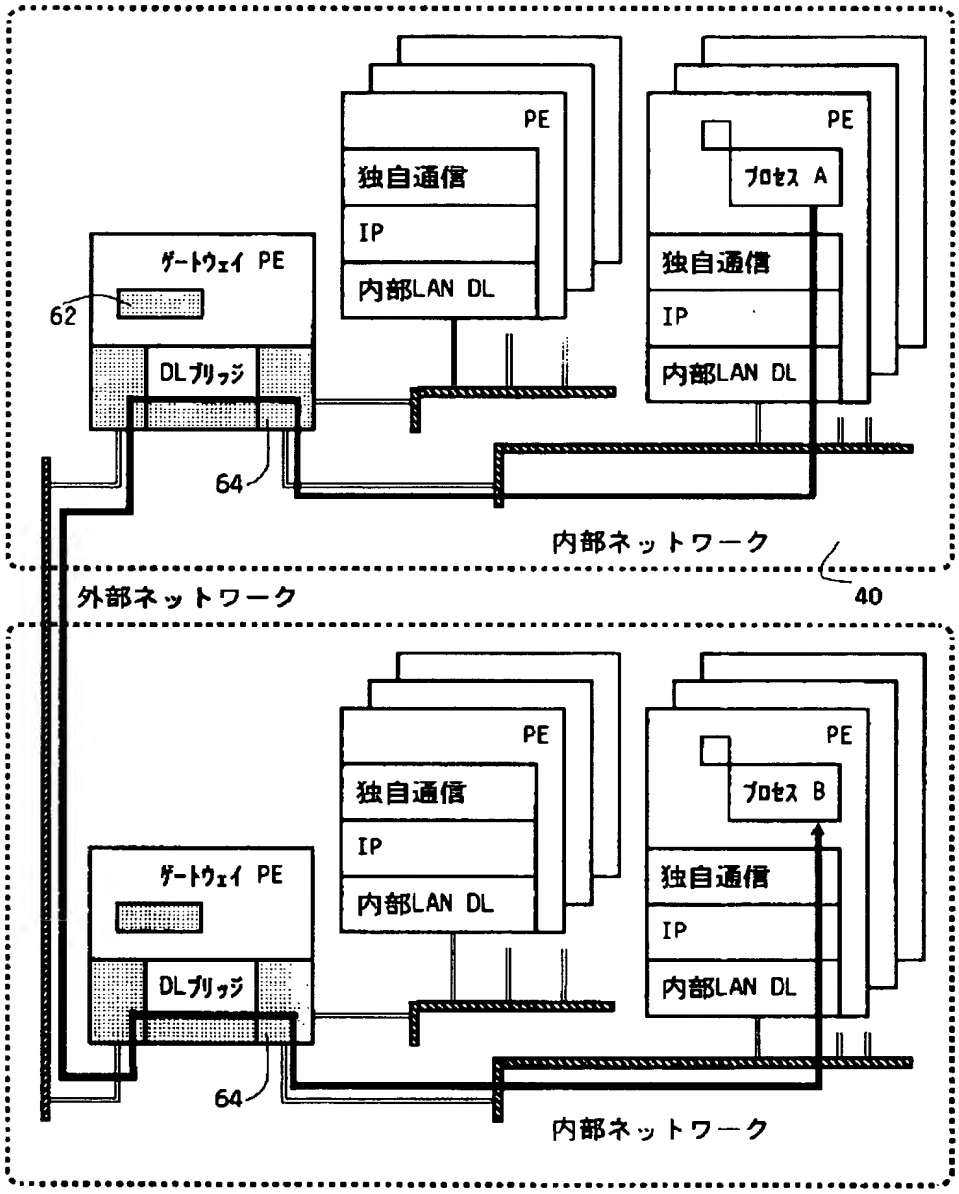
【図１５】

ハブを用いた通信宛先管理システム構成図



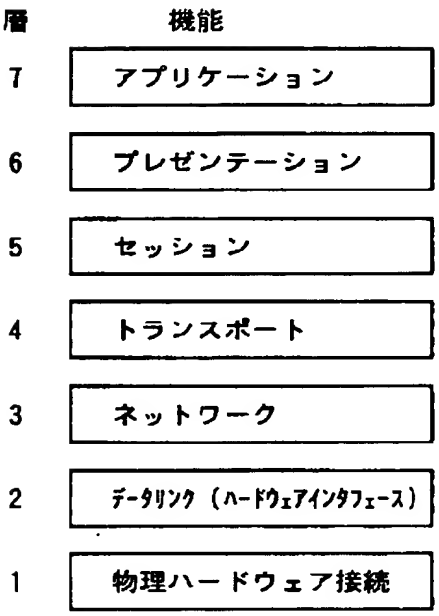
【図１６】

大規模分散処理における通信宛先管理システム構成図



【図１７】

ＩＳＯの７層参照モデル



国際標準化機構
(International Organization for Standardization (ISO))

【図１９】

従来のプロセス間通信